|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Working Title of the Project:** | | **Information Extraction using Large Language Models** | | |
| **Project Site / Location** | | Chennai | | |
| **Name and address of the company / organisation**  **(Applicable for projects with industry or industry support)** | | SRM IST, Kattankulathur, Chengalpattu District-603203 | | |
| **Supervision Team** | | | | |
|  | **Supervisor** | | **Co-Supervisor** | **External Supervisor**  **(If applicable)** |
| **Name** | Dr. Maivizhi R | |  |  |
| **Designation** | Assistant Professor | |  |  |
| **Department** | CINTEL | |  |  |
| **Campus** | Kattankulathur | |  |  |
| **Telephone** | 9655554718 | |  |  |
| **E-mail** | maivizhr@srmist.edu.in | |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name of student** | **Register Number** | **Department** | **Mobile Number** | **Email ID** |
| Dibyajyoti Ganguly | RA2011026010122 | CINTEL | 8337090777 | dg7255@srmist.edu.in |
| Tarun Negi | RA2011026010132 | CINTEL | 7985857522 | tn9791@srmist.edu.in |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Degree/ program** | B.Tech | **Specialisation** | Computer Science& Engineering  **Project Batch ID**  **CIN\_238** |
| **Academic Year** | 2023-2024 (Even) | **Semester** | 8 |
| **Course Code** | **18CSP109L** | **Major Project** | Information Extraction using Large Language Models |
| Mission Statement  To develop a system that accurately extracts key information from documents, including facts, quotes, and contextual details, to ensure the reliability of the extracted content. | | | |
|  | | | |
| **Problem (or) Product Description:** | | | |
| **Efficiently extracting information from documents remains a critical challenge in various domains. As we witness an exponential growth in online sources, the need for efficient and intelligent tools to distill and comprehend this wealth of information becomes more pronounced. The traditional methods of research often fall short in managing the sheer volume of data available, leading to information overload and reduced productivity. Our proposed work leverages the powerful combination of Langchain and OpenAI.** | | | |
|  | | | |
| **Assumptions : accurate data, with a chunk size of 1000, the system achieved higher precision but lower recall compared to 500.** | | | |
| **Constraints : only supported document file format is PDF, system accepts only textual data, not images.** | | | |
| **Stakeholders : Journalists, Equity Research Analysts, Large-scale businesses, Educational institutions** | | | |

**Division of work and contributors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time period** | | **Activities or components of the project** | **Name/Register Number of the Individual Contributor** | **Names/Register Number of the Joint Contributors** |
| **From Date** | **To Date** |
| 4/01/24 | 6/01/24 | Working on project problem statement | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
| 12/01/24 | 13/01/24 | Working on base paper and data preprocessing | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
| 15/02/24 | 16/02/24 | Preparing Literature survey | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
| 20/02/24 | 23/02/24 | Working on code implementation | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
| 4/03/24 | 08/03/24 | Finishing up research paper and comparing the results | TARUN NEGI | DIBLYAJYOTI GANGULY |
| 2/04/24 | 4/04/24 | Working on project problem statement | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
| 22/04/23 | 26/04/24 | Working on project report | TARUN NEGI | DIBLYAJYOTI GANGULY |
|  |  |  |  |  |
|  |  |  |  |  |

**Division of work and contributors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time period** | | **Activities or components of the project** | **Name/Register Number of the Individual Contributor** | **Names/Register Number of the Joint Contributors** |
| **From Date** | **To Date** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Summary record of major progress meetings with supervisors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summary record of major progress meetings with supervisors** | |  | **Working title of dissertation/research project:** | |
| **Meeting date & supervisors present** | **Progress since last meeting** | **Agreed programme of work and target dates** | **Other issues, e.g. facilities, supervision, training needs, etc.** | **Date of next meeting** |
| 19/01/24 | Started working on the research paper | Started searching for related research papers on IEEE |  | 15/02/24 |
| 15/02/24 | Literature survey for 15 IEEE papers | Finalising base paper and studying it thoroughly |  | 4/03/24 |
| 4/03/24 | Code implementation | Working on 75% implementation of the code |  | 21/03/24 |

**Summary record of major progress meetings with supervisors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summary record of major progress meetings with supervisors** | |  | **Working title of dissertation/research project:** | |
| **Meeting date & supervisors present** | **Progress since last meeting** | **Agreed programme of work and target dates** | **Other issues, e.g. facilities, supervision, training needs, etc.** | **Date of next meeting** |
| 21/03/24 | Full code implementation | Working demo of the project and comparison with other models and report |  | 22/04/24 |
| 22/04/24 | Report Completion and plagiarism check |  |  |  |
|  |  |  |  |  |

**Summary record of major progress meetings with supervisors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summary record of major progress meetings with supervisors** | |  | **Working title of dissertation/research project:** | |
| **Meeting date & supervisors present** | **Progress since last meeting** | **Agreed programme of work and target dates** | **Other issues, e.g. facilities, supervision, training needs, etc.** | **Date of next meeting** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Worksheet / Data collection / Observation etc**

**Journal Publication Details**